5К

УДК 004.85:004.81

Н.В. Крачковский
Донецкий национальный университет, Украина
Украина, 83000, г. Донецк, пр. Театральный, 13

О модификации метода обучения с подкреплением на основе моделей когнитивной психологии

M.V. Krachkovsky
Donetsk National University, Ukraine
Ukraine, 83000, c. Donetsk, Teatralnyi av., 13

About Reinforcement Learning Method Modification Based on Cognitive Psychology Models

М.В. Крачковський
Донецький національний університет, Україна
Україна, 83000, м. Донецьк, пр. Театральний, 13

Про модифікацію метода навчання з підкріпленням на основі моделей когнітивної психології

В статье рассматривается задача обучения системы мотивированного контекстного ситуационного управления. Описаны модели структурных изменений множества агентов при обучении, показана формализация этапа формирования прототипов ситуации и реакции, а также контекстной связи. Проведены компьютерные эксперименты, демонстрирующие процесс обучения.
Ключевые слова: ситуационное управление, обучение с подкреплением, когнитивная психология.

We consider the problem of learning of motivated context situational control system. There are described models of structural changes in the set of agents at training, shown formalization of situation and reaction prototype as well as the context link. The computer experiments demonstrating the learning process are made.
Key words: situational control, reinforcement learning, cognitive psychology.

У статті розглядається задача навчання системи мотивованого контекстного ситуаційного керування. Описані моделі структурних змін множини агентів під час навчання, показана формалізація етапу формування прототипів ситуації та реакції, а також контекстного зв’язку. Проведені комп’ютерні експерименти, що демонструють процес навчання.
Ключові слова: ситуаційне керування, навчання з підкріпленням, когнітивна психологія.

Введение

В статье рассматривается задача обучения поведению сложных робототехнических комплексов, которые могут использоваться либо для снижения производственных затрат, либо в случаях, когда непосредственное управление человеком затруднено. Поведение, которое должна демонстрировать система, заранее запрограммировать затруднительно в условиях отсутствия полной информации на этом этапе (функционирование в открытой среде). Для управления такими комплексами применяются ситуационные системы управления [1]. Возникновение новых требований к поведению системы в процессе её

функционирования требует обучения этой системы. Известные подходы к обучению, в основном, базируются на моделях искусственных нейронных сетей [2], [3], поведенческих сетей [4], развивающегося интеллекта [5].

Рассматриваемая модель ситуационного управления [1], [6], модифицирована на основе данных когнитивной психологии [7], [8], которая подобно человеку и высокоорганизованным животным хранит в памяти не набор прототипов «ситуация-действие», характерный для классических систем ситуационного управления, а прототип последовательностей действий, названные скриптами. Особенность модели контекстного ситуационного управления выражена в структуре правил в виде односторонней зависимости правил (1), если ввести понятие контекста.

$$\Pi_i : \text{ЕСЛИ } \{ cont_{j,i} ,\ S \subset \hat{S}_i ,\ M \} \text{ ТО } \{ u_i ,\ cont_{i,l} \}, \qquad (1)$$

где M – мотив, S – текущая ситуация, $\hat{S}$ – эталонная ситуация-прототип, u – управляющее воздействие, $cont_{j,i}$ – контекстная связь между правилами Π_j и Π_i.

Схематично организацию системы мотивированного контекстного ситуационного управления можно представить в виде, показанном на рис. 1. Она включает следующие компоненты: множество сенсоров $SN = \{sn_i\}_{i=1}^{ns}$, множество ситуационных агентов $CA = \{CA_j\}_{j=1}^{na}$, множество эффекторов $U = \{u_k\}_{k=1}^{nu}$, множество мотивов $M = \{m_l\}_{l=1}^{nm}$. Поведение системы определяется взаимодействием агентов с окружением: ситуацией, формирующей значения нечётких характеристик сенсорных элементов и мотивов.

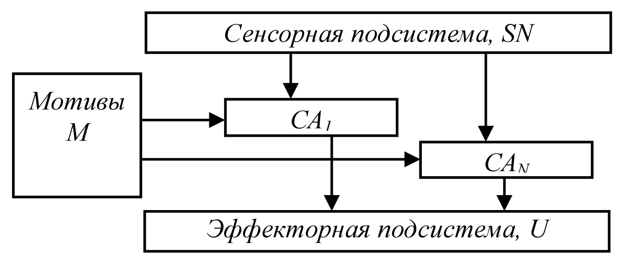


Рисунок 1 – Общая схема системы управления

Ситуационный агент представляет упорядоченное контекстом множество ситуационных элементов $\{ce_i\}_{i=0}^{n}$, как показано на рис. 2. Каждый ситуационный элемент описывается правилом (1).

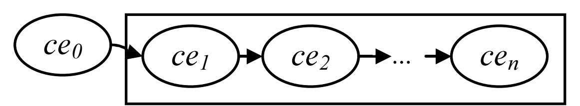


Рисунок 2 – Схематичное строение ситуационного агента

В статье рассматривается задача обучения такой системы управления.

Постановка задачи. Система мотивированного контекстного управления [9], как и традиционная система ситуационного управления, базируется на множестве контекстно-зависимых правил.

Управление рассматривается как многошаговый дискретный процесс в моменты времени t, $t+T$, $t+2T$, …, $t+kT$,... Последовательность этапов одного шага управления представлена на рис. 3: сформированные физическим датчиком значения фаззифицируются в виде нечётких характеристик элементарных сенсоров, которые формируют сенсорную память; на основании сравнения текущей ситуации и прототипов

ситуации из прототипной памяти, формируется нечёткие характеристики прототипов реакций эффекторной памяти. Последний этап заключается в преобразовании прототипа реакции в непосредственную реакцию – дефаззифицированные значения подаются исполнительному механизму.

Прототипная память системы представлена набором ситуационных элементов, сгруппированных в ситуационные агенты. Отдельный ситуационный агент представляет некоторое отдельное законченное действие – фрагмент поведения.

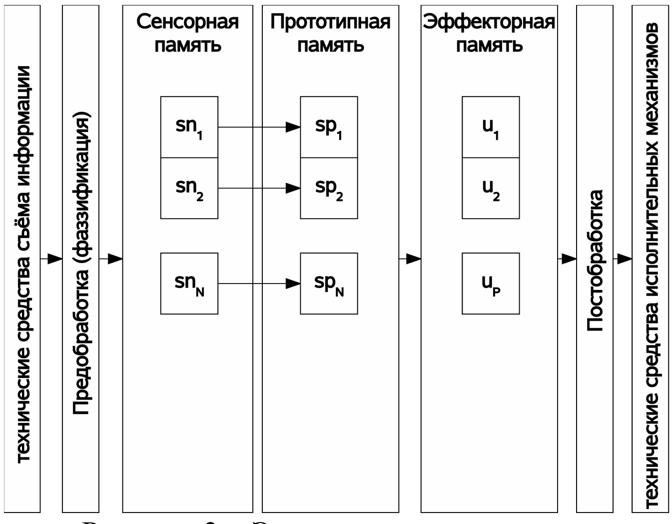


Рисунок 3 – Этапы шага управления

Для появления у системы нового поведения требуется создание нового ситуационного агента или модификация одного из существующих ситуационных агентов. Пополнение множества ситуационных агентов выполняется методом обучения. В качестве исходного метода обучения, который развивается применительно к рассматриваемому классу систем, применяется обучение с подкреплением [10].

В статье излагается модель и метод обучения с подкреплением для автономного формирования новых ситуационных агентов из нескольких ситуационных элементов, связанных в контекстную цепочку. Метод базируется на обобщении теорий научения (Э. Торндайка, Б. Скиннера, И. Павлова [11]), изученных в когнитивной психологии.

Формальная модель управления

Ситуационный элемент ce контекстной цепочки (рис. 2), характеризуется:

1) нечётким прототипом ситуации – $\hat{S}$;
2) нечётким прототипом управления – $\hat{R}$;
3) контекстной связью – K;
4) мотивированной связью – M.

Каждая из данных характеристик представляет собой множество нечётких характеристик [12] вида (2):

$$\hat{A} = \left\{ \underset{\sim}{A}_i = \left\{ x \mid \mu_{\underset{\sim}{A}_i}(x) \right\} \right\}, x \in [-1, +1], \quad \mu_{\underset{\sim}{A}_i}(x) = 2 \cdot exp\left(-\frac{\left(x - \alpha_{\underset{\sim}{A}_i} \right)^2}{2\beta^2_{\underset{\sim}{A}_i}} \right) - 1, \qquad (2)$$

Из 4-х вариантов рассмотренной концептуальной модели [13] в статье рассматривается задача обучения, сводящаяся к формированию нового ситуационного агента, его расширения и модификации контекстной связи.

Создание нового ситуационного агента рассматривается как многоэтапный процесс обучения, на каждом этапе которого формируется ситуационный элемент путём нахождения вышеперечисленных характеристик: прототипов ситуации ($\hat{S}$) и управления ($\hat{R}$); нечётких характеристик мотива (M) и контекстной связи (K).

Первый шаг каждого этапа обучения начинается с обработки информации для выделенного не специфицированного «пустого» ситуационного элемента, который будет служить базой для образования нового элемента. Данный элемент обладает потенциальными связями со всеми существующими компонентами: контекстные с агентами; информационные с сенсорами и управлением; и связи с мотивами. Изначально эти связи имеют нейтральные значения нечётких характеристик. Структура ситуационного элемента и его потенциальные связи показаны на рис. 4.

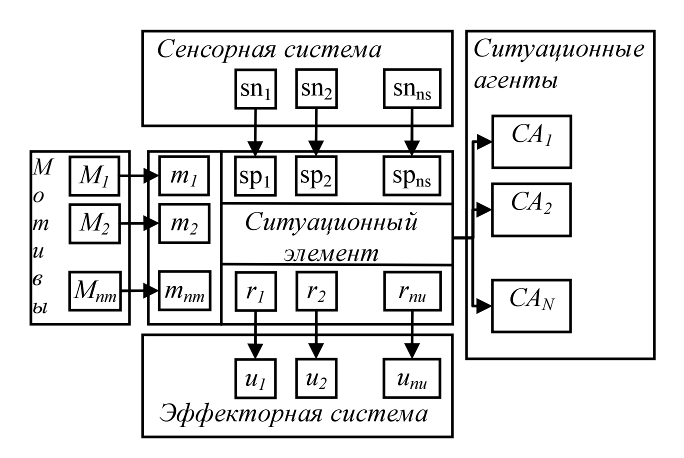


Рисунок 4 – Ситуационный элемент

На последующих i-х шагах обучение происходит в моменты времени kT на основе подкрепления: произошло изменение (падение) нечёткой характеристики активности мотива. Значения перечисленных характеристик ситуационного элемента в моменты времени kT находятся по модели обучения F на основании значений этих характеристик в предыдущий момент времени, а также вектора активности сенсоров (MS), действий (MR) и мотива (MM).

$$\left\langle \hat{S}, \hat{R}, M, K \right\rangle_{kT} = F\left(\left\langle \hat{S}, \hat{R}, M, K \right\rangle_{(k-1)T}, MS, MR, MM \right) \quad (3)$$

где $MS = \langle S((k-j)T) \rangle_{j=0}^{d}$, $MR = \langle R((k-1)T) \rangle_{j=0}^{d}$, $MM = \langle M((k-j)T) \rangle_{j=0}^{d}$,

$$S(kT) = \left\{ \underset{\sim}{A}_i^{sn}(kT) \right\}_{i=1}^{ns}, \; R(kT) = \left\{ \underset{\sim}{A}_i^{u}(kT) \right\}_{i=1}^{nu}, \; M(kT) = \left\{ \underset{\sim}{A}_i^{M}(kT) \right\}_{i=1}^{nm}.$$

Ниже рассматривается формализация процедуры формирования прототипов $\hat{S}$ и $\hat{R}$ и контекста модели F обучения (3).

Концептуальная модель обучения

При формализации механизма связанного с изменением базы знаний объекта управления, принято во внимание следующее: обучение происходит в том случае, когда имеет место фактор «неожиданности», так если для активного мотива и сложившейся ситуации существует агент, выполнение функции которого приводит к погашению мотива, то новых знаний система не приобретает.

В случае отсутствия такого агента либо реагирование какого-либо другого агента, не приводящее к погашению мотива, означает, что существующие схемы неэффективны и требуется обучение. В таком случае запускается («включается») механизм обучения. Он использует информацию о ситуации, из которой произошёл переход к какой-либо известной ранее. В данном случае должны закрепляться: предыдущая ситуация и выполненное действие в виде ситуационного элемента, а также контекстная связь между данным элементом и существующим ситуационным агентом, которая будет определять ожидаемость погашения мотива.

Возможно ещё, когда случайно выработанное управление привело к погашению мотива, – подкреплению. В таком случае полученный ситуационный элемент (СЭ) образует новый ситуационный агент, состоящий из одного ситуационного элемента. В этих двух случаях идёт образование нового элемента.

Приведённые выше рассуждения являются обобщением известных теорий научения из физиологии и когнитивной психологии [11], а именно теорий Э.Л. Торндайка, К.Л. Халла, Э.Ч. Толмена, А. Бандуры. Анализ этих теорий и вышеприведённые рассуждения позволили обобщить и выделить 4 варианта обучения:

1. Изменение контекстной связи между ситуационными агентами.

Это происходит в случае, когда ситуация, полученная в результате выполнения функции CA_i, сопоставима с прототипом ситуации, необходимым для активации другого CA_j. Многократное повторение такой последовательности с последующим подкреплением (ослабление мотива) приводит к усилению контекстной связи $CA_i \rightarrow CA_j$ и в дальнейшем даже при значительном отклонении ситуации контекстная связь может обеспечить активацию ситуационного агента CA_j.

2. Образование нового ситуационного элемента.

Если в процессе случайного применения управления образовалась ситуация, подходящая под прототип первого ситуационного элемента агента CA_k, поведение согласно которому привёло к погашению мотива, то активный мотив, исходная ситуация и выработанное поиском действие становятся мотивом, прототипами ситуации и реакции соответственно нового ситуационного элемента. Также образуется контекстная связь между вновь созданным СЭ и CA_k.

3. Образование нового ситуационного агента.

Возможно, что в процессе принятия управления, привело к погашению мотива – подкреплению. В таком случае полученный СЭ образует новый ситуационный агент, состоящий из одного ситуационного элемента.

4. Изменение прототипа существующего ситуационного элемента

Происходит, если текущая ситуация была близка к прототипу некоторого существующего ситуационного элемента, и было получено подкрепление.

На рис. 5 приведено 3 варианта обучения. На рис. 5 а) показано изменение контекстной связи (сплошная стрелка) между ситуационными агентами CA_i и CA_j. При этом связь формируется односторонняя – в том же порядке, в котором происходит

выполнение агентов. На рис. 5 б) показано формирование ситуационного агента CA_k путём внесения в него нового ситуационного элемента. На рис. 5 в) приведён новый ситуационный агент CA_z, сформированный на базе одного элемента.

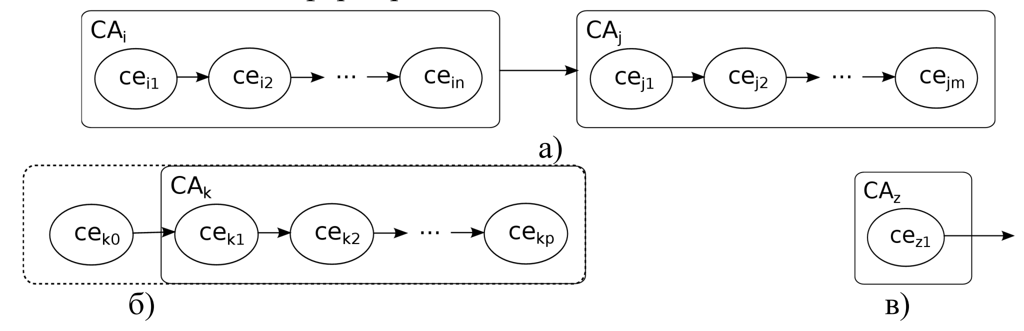


Рисунок 5 – Изменения структуры системы, вносимые обучением

В первом варианте механизм обучения формирует нечёткое множество контекстной связи $co_{i,n,j,1}$, которая влияет на активность суммарного контекстного входа $\underset{\sim}{A}_{(k-1)T}\left(co_{j,1}\right)$, входящей в модель управления [9] при расчёте активности ситуационного элемента $ce_{j,1}$.

Во втором варианте механизма обучения формируются нечёткие множества прототипа ситуации $\hat{S}_{k,0}$, входящего в расчёт активности ситуационного элемента, и прототипа реакции $\hat{R}_{k,0}$, входящего в расчёт нечёткой активности эффекторов u_x в модели управления [12]. Кроме этого также формируется контекстная связь $co_{k,0,k,1}$.

В третьем варианте, как и во втором, происходит формирование нечётких множеств прототипа ситуации $\hat{S}_{z,1}$ и реакции $\hat{R}_{z,1}$. Однако контекстная связь формируется между контекстным элементом и подкреплением и описывает ожидание подкрепления в случае выполнения данного ситуационного агента. Данная контекстная связь используется при управлении для выбора подходящего агента, в случае наличия альтернатив, а также при дальнейшем обучении.

Значения всех сформированных нечётких множеств в каждом из вариантов 1 – 3 зависят от времени, прошедшего между предъявлением стимула и изменением активности ожидаемых значений мотивов (стимулом потребности), величины изменения активности мотива (полученного подкрепления).

Формальная модель обучения

Прототип ситуации представлен множеством нечётких характеристик элементов $\{sp_i\}$, соответствующих сенсорам $\{sn_i\}$ сенсорной системы. Формирование прототипа ситуации, который представлен в виде множества (2), в kT момент времени вычисляется согласно выражению (4).

$$\underset{\sim}{A}_i^{sp}(kT) = arg \min_{j=0,d} \beta_{\underset{\sim}{A'}_i}\left((k-j)T\right), \tag{4}$$

где $\underset{\sim}{A'}_i\left((k-j)T\right)$ – расчётная нечёткая характеристика элемента sp_i, модифицированная с учётом влияния эффективности обучения и величины подкрепления относительно момента времени $(k-j)T$; d – глубина сенсорной памяти.

Формализация $\underset{\sim}{A'_i}((k-j)T)$ приведена на (5).

$$\underset{\sim}{A'_i}((k-j)T)=\left\{x\middle|\mu(x)=exp\left(-\frac{\left(x-\alpha_{A'_i((k-j)T)}\right)^2}{2\cdot\left(\beta_{A'_i((k-j)T)}\right)^2}\right)\right\}, \tag{5}$$

где $\alpha_{\underset{\sim}{A'_i}((k-j)T)}=(1-q)\cdot\alpha_{\underset{\sim}{A}_i^{sp}(kT)}+q\cdot\alpha_{\underset{\sim}{A}_i^{sn}((k-j)T)}$; $\beta_{\underset{\sim}{A'_i}((k-j)T)}=(1-q)\cdot\beta_{\underset{\sim}{A}_i^{sp}(kT)}+q\cdot\beta_2$;

$$\beta_2=\frac{4\beta_{\underset{\sim}{A}_i^{sn}}((k-j)T)}{I_{m_S}(jT)\left(\alpha_{\underset{\sim}{Q}^{\overline{M}}(kT)}+1\right)\left(\alpha_{\underset{\sim}{A}_i^{\Delta sn}((k-j)T)}+1\right)}+\frac{\beta_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}+\beta_{\underset{\sim}{A}_i^{\Delta sn}((k-j)T)}}{I_{(m_S)}(jT)};\ q=\varphi\cdot e^{-\beta_2};$$

φ – параметр скорости обучения;

$\underset{\sim}{A}_i^{\Delta sn}((k-j)T)$ – нечёткая характеристика скорости изменения сенсора sn_i в момент времени $(k-j)T$;

$\underset{\sim}{Q}_j^{\overline{M}}(kT)$ – нечёткая характеристика подкрепления.

По данным когнитивной психологии [11] процесс научения происходит с различной эффективностью, которая определяется такими параметрами, как время между предъявлением стимула, совершённой реакцией и полученным подкреплением. Предлагается эту зависимость представить в виде (6).

$$I_m(x)=\frac{x}{m}\cdot e^{1-\frac{x}{m}},x\geq 0 \tag{6}$$

где x – время от предъявления стимула до подкрепления, для $x<0$ можно считать значение равным 0; m – параметр, задающий значение оптимального времени.

Вторым фактором, влияющим на эффективность научения, является величина подкрепления – явились ли последствия действия полезными для объекта. Формализация представлена ниже (7).

$$\underset{\sim}{Q}_j^{\overline{M}}(kT)=\left\{x\mid\mu(x)=exp\left(-\frac{\left(x-\alpha_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}\right)^2}{2\cdot\beta^2_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}}\right)\right\}, \tag{7}$$

$$\alpha_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}=m\left(\alpha_{\underset{\sim}{A}^{\overline{M}}((k-j)T)}-\alpha_{\underset{\sim}{A}^{\overline{M}}(kT)}\right),\ \beta_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}=\beta_{\underset{\sim}{A}^{\overline{M}}((k-j)T)}-\beta_{\underset{\sim}{A}^{\overline{M}}(kT)},$$

$$m(x)=\frac{x}{2}\cdot\left|\frac{x}{2}\right|^{\gamma},-1<\gamma\leq 0,$$

Параметр γ влияет на эффективность обучения при малых изменениях мотива.

Реальное изменение мотива может быть слишком отложено во времени, чтобы привести к обусловливанию, поэтому в работах физиологов вводилось понятие стимула потребности [11]. Он формализован в данной работе при описании подкрепления как суммарный мотив, который определяется на основании реального и фантомного мотивов. Под фантомным мотивом понимается ожидаемое значение мотива и является основой упреждения системы (8).

$$\underset{\sim}{A}_i^{\overline{M}}(kT) = \underset{\sim}{A}_i^{M}(kT) \oplus \underset{\sim}{A}_i^{M'}(kT), \tag{8}$$

где M' – фантомная активация мотива, $\oplus$ – операция нечёткого накопления.

Данная форма позволяет описать такие явления как стимул потребности, научение при отсутствии реального мотива (например, ситуации опасности), а также отсутствие научения в случае эффективности существующих реакций.

Прототип реакции, аналогично прототипу ситуации, представлен множеством вида (2). Элемент данного множества определяется по формуле (9).

$$\underset{\sim}{A}_i^{r}(kT) = arg \min_{j=0,d} \beta_{\underset{\sim}{A'}^R((k-j)T)}, \tag{9}$$

где $\underset{\sim}{A'}^R((k-j)T)$ – расчётная нечёткая характеристика управления r_i, модифицированная с учётом влияния эффективности обучения и величины подкрепления на момент времени $(k-j)T$, значение характеристики находится аналогично (5).

Параметр эффективности обучения (6) для реакции должен быть меньше, чем параметр для ситуации, так как реакция выполняется с некоторой задержкой после предъявления стимула.

При формировании контекстной связи (10) между ситуационными элементами ce_h и ce_k величина нечёткой характеристики этой связи будет изменяться в соответствии с полученным подкреплением (7).

$$\alpha_{\underset{\sim}{A}_{k,h}^{co}(kT)} = (1-\varphi)\cdot\alpha_{\underset{\sim}{A}_{k,h}^{co}((k-1)T)} + \varphi\cdot\alpha_{\underset{\sim}{Q}_j^{\overline{M}}(kT)}. \tag{10}$$

Компьютерный эксперимент

Компьютерный эксперимент проводился на комплексе, состоящем из робота LEGO MINDSTORMS NXT, который удалённо управляется компьютером при помощи Bluetooth. Робот имеет два независимо управляемых колеса. Третье колесо пассивное и обеспечивает устойчивость. Из датчиков робота использованы датчик соприкосновения и расстояния, оба направлены вперёд. Целью эксперимента было обучение робота поведению, которое позволяет избежать столкновения с препятствием.

В качестве примера рассмотрим обучение, которое будет выражаться в образовании ситуационного агента, вначале состоящего из одного ситуационного элемента. Для формирования данного ситуационного элемента необходимы: мотив, прототип ситуации и прототип реакции. Контекст в данном случае будет «нулевым».

Информация от двух датчиков гранулирована так, что по показаниям датчиков формируются нечёткие характеристики 28 элементарных сенсоров: 2 сенсора для датчика со-

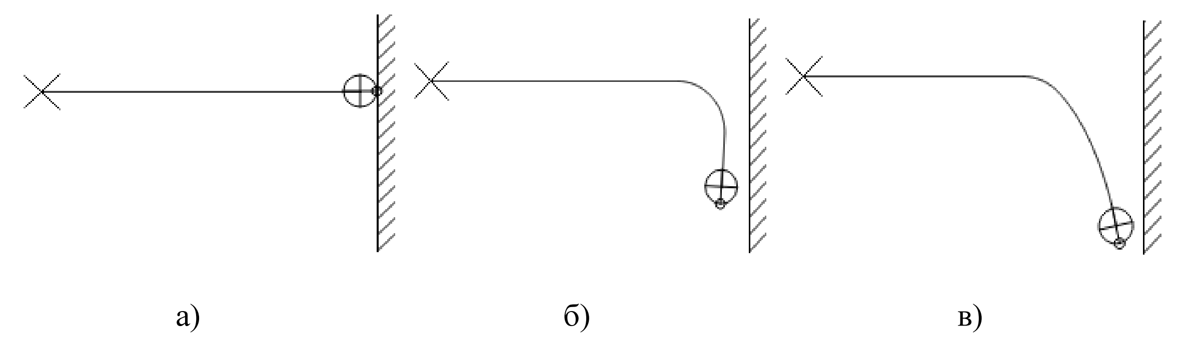
прикосновения (snt_0, snt_1), 20 для датчика расстояния для разного уровня детализации (от 2 до 6 сенсоров на область детектирования датчика: $snd_{i,j}, i = \overline{1,5}, j = \overline{0,i}$), а также по 3 сенсора на каждое колесо ($snr_{i,j}, i = \overline{0,1}, j = \overline{-1,1}$). Подробнее датчики описаны в [9].

В качестве мотива выбран мотив самосохранения робота, который основан на snt_1: если сработал датчик, то есть угроза столкновения. В случае столкновения робота с препятствием, обучение избеганию будет происходить в два этапа: формирование знания о столкновении и собственно обучение избеганию. В случае обучения сложному поведению эти этапы будут повторяться.

Обучение было выполнено по следующей схеме. Робот движется по прямой к стенке со средней скоростью. При столкновении со стенкой возрастает активность мотива самосохранения. Ситуация соответствует прототипу изначально закреплённого агента CA_1, выдаёт управление в соответствии с прототипом реакции – робот останавливается и отъезжает от стенки. До обучения траектория движения робота показана на рис. 6 а) и представляет собой горизонтальную линию. Вертикальной линией показана стена.

В момент активизации мотива самосохранения происходит обучение упреждающей ситуации – формируется прототип, описывающий малое расстояние до препятствия и движение вперёд. В процессе компьютерного эксперимента на основании нескольких столкновений данный прототип закрепляется и начинает активизировать фантомный мотив самосохранения до столкновения.

Затем, при приближении робота к стене, подаётся команда поворота вправо, которая позволяет избежать столкновения со стеной, траектория показана на рис. 6 б), мотив не активируется, а значит, активность $\overline{M}$ падает, что соответствует подкреплению. В этом случае происходит обучение системы: формируется прототип реакции (9). В качестве мотива используется мотив, который был погашен — мотив самосохранения. «Пустой» ситуационный элемент специфицируется. При повторении ситуации, в которой ожидается столкновение, и поворот вправо позволяет избежать его, снова происходит обучение – прототипы ситуационного элемента модифицируются и закрепляются. На рис. 6 в) показана траектория, выработанная управлением вновь сформированного ситуационного элемента после серии экспериментов обучения.



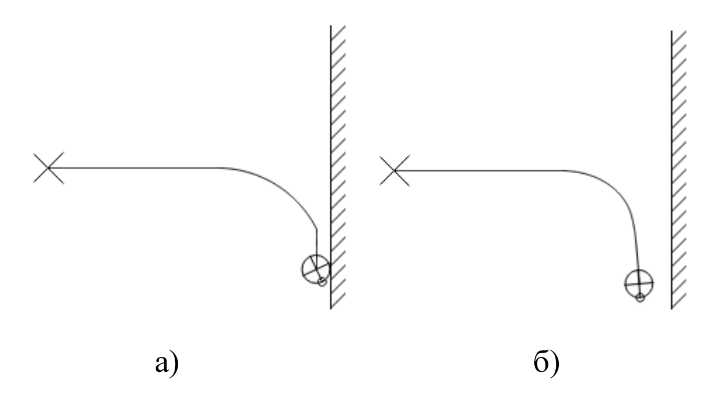
а) б) в)

Рисунок 6 – Поведение робота при столкновении со стенкой: до обучения (а), эталонная реакция (б) и после обучения (в)

Таким образом, закрепление успешной реакции позволяет системе избегать столкновения с препятствием в дальнейшем без необходимости поиска.

На следующем этапе аналогичным методом формируется второй ситуационный элемент агента, для ситуации, когда робот движется с высокой скоростью и не успевает повернуть. Данная ситуация представлена на рис. 7 а).

Обучение в данном случае будет проведено в три этапа. На первом этапе будет произведена дифференцировка прототипа ситуации ранее описанного элемента: высокое значение начальной скорости не является подходящим, так как не получено подкрепление.



а) б)

Рисунок 7 – Траектория движения при высокой начальной скорости

На втором этапе, который будет происходить частично параллельно с первым, происходит выделение нового ситуационного элемента. Данный этап абсолютно аналогичен рассмотренному ранее процессу, и полученный прототип ситуации данного элемента также значительно отличается только в сенсорах $srd_{i,j}$.

Третий этап заключается в поиске подходящей реакции, которая позволит снизить активность мотива самосохранения. В нашем случае из возможных найденных решений рассмотрим реакцию снижения скорости движения. В этом случае, непосредственно данная реакция не приводит к подкреплению, однако, возникает ситуация, которая соответствует уже известному прототипу. В результате применяется первый ситуационный элемент, который и является подкрепляющим стимулом для второго. Данные ситуационные элементы связываются контекстной связью и второй ситуационный элемент теперь является частью ситуационного агента. Результат работы агента из двух ситуационных элементов показан на рис. 7 б).

Выводы

Рассмотрены варианты механизма обучения обобщённого ситуационного управления, которые приводят к структурным изменениям системы управления. Формальная модель описывает зависимость начальных значений сформированных прототипов ситуации и реакции ситуационного элемента от других известных величин.

При дальнейшем функционировании системы может также происходить обучение, влияние которого отражается не на структуре системы, а на значении прототипов существующих элементов и контекстных связей между ситуационными элементами.

Предложен новый подход к обучению в ситуационных системах управления, отличающийся от известных, базирующихся на нейронных сетях и нечетких системах, тем, что в нём формализованы теории научения, освещённые в когнитивной психологии. Рассмотрена формализация процедуры формирования прототипов ситуации и управления в задаче самообучения.

Литература

1. Поспелов Д. А. Ситуационное управление: Теория и практика [текст] / Д. А. Поспелов – М. : Наука. – Гл. ред. физ.-мат. Лит., 1986. – 288 с.
2. Терехов В.А. Нейросетевые системы управления [текст] / В.А. Терехов, Д.В. Ефимов, И.Ю. Тюкин – М.: Высш. шк., 2002.
3. Tan A.-H. Intelligence through interaction: towards a unified theory for learning [текст] / A.-H. Tan, G.A. Carpenter, S. Grossberg. – Advances in neural networks. – 2007. – № 1. – P. 1094-1103.
4. Maes P. Learning to Coordinate Behaviors [текст] / P. Maes, P. Brooks – AAAI Press/MIT Press – Proceedings of the Eighth National Conference on Artificial Intelligence, 1990. – P.796-802.
5. Meng Y. Bio-Inspired Self-Orginizing Robotic Systems [текст] / Yan Meng, Yaochu Jin. – Springer-Verlag Berling Heidelberg – 2011. – 273 p.
6. Мелихов А.Н. Ситуационные советующие системы с нечеткой логикой [текст] / Мелихов А.Н., Берштейн Л.Е., Коровин С.Д. – М. : Наука, 1990.
7. Солсо Р. Когнитивная психология [текст] / Р. Солсо. – СПб. : Питер, 2002. – 592 с.
8. Андерсон Дж. Р. Когнитивная психология [текст] / Дж. Р. Андерсон. – СПб. : Питер, 2002. – 496 с.
9. Каргин А. А. Об одной модели ситуационного управления подвижным роботом [текст] / А. А. Каргин, Н. В. Крачковский// Інформаційно-керуючі системи на залізничному транспорті. – 2011. – № 4(89). – С. 12-17.
10. Саттон Р. Обучение с подкреплением [текст] / Р. Саттон, Э. Барто. – СПб. : Бином, 2011. – 399 с.
11. Хегенхан Б. Теории научения [текст] / Б. Хегенхан, М. Олсон. ; пер. на русс. яз. ЗАО Издательский дом «Питер». – [6-е изд.]. – СПб. : Питер, 2004. – 474 с. : ил. – (Серия «Мастера психологии»).
12. Каргин А. А. Введение в интеллектуальные машины. Книга 1. Интеллектуальные регуляторы [текст] / А. А. Каргин. – Донецк : Норд-Пресс, ДонНУ, 2010. – 526с.
13. Каргин А. А. Модели обучения системы мотивированного контекстного ситуационного управления [текст] / А. А. Каргин, Н. В. Крачковский // Вісник ХНТУ. – 2012. – №1(44). – С.257-260

Literatura

1. Pospelov D. A. Situational control: Theory and practice [text] / D. A. Pospelov – Moscow: Nauka. – Main Publ. Phys.-Math. Lit., 1986. – 288 p.
2. Terekhov V. A. Neural network control system [text] / V. A. Tetekhov, D. V. Yefimov, I. Yu. Tiukin – Moscow: High School, 2002.
3. Tan A.-H. Intelligence through interaction: towards a unified theory for learning [текст] / A.-H. Tan, G.A. Carpenter, S. Grossberg – Advances in neural networks. – N. 1., 2007. – P.1094–1103.
4. Maes P. Learning to Coordinate Behaviors [текст] / P. Maes, P. Brooks – AAAI Press/MIT Press – Proceedings of the Eighth National Conference on Artificial Intelligence, 1990, P.796-802.
5. Meng Y. Bio-Inspired Self-Orginizing Robotic Systems [текст] / Yan Meng, Yaochu Jin – Springer-Verlag Berling Heidelberg – 2011. – 273p. – ISBN 978-3-642-20759-4.
6. Melikhov A. N. Situational advise systems with fuzzy logic [text] / A. N. Melikhov, L. Ye. Bershtein, S. D. Korovin – Moscow: Nauka, 1990.
7. Solse R. Cognitive Psychology [text] / R. Solso – St. Petersburg: Piter, 2002. – 592 p.
8. Anderson J. R. Cognitive psychology [text] / J. R. Anderson – St. Petersburg: Piter, 2002. – 496 p.
9. Kargin A. A. About the model of situational control of mobile robot [text] / A. A. Kargin, M. V. Krachkovsky – Kharkiv: Science-technical magazine «Informatsiino-keruiuchi systemy na zaliznychnomu transporti» – 2011. – №4(89).-P.12-17
10. Satton R. Reinforcement learning [text] / R. Satton, E. Barto – St. Petersburg: Binom, 2011, - 399 p.
11. Hergenhahn B. Introduction to the Theories of Learning [text] / B. Hergenhahn, M. Hergenhahn; Russian Translation CJSC Publishing House «Piter». – [6th issue]. – St. Petersburg: Piter, 2004. – 474 p. – («Psychology Masters» series). – ISBN 5-94723-033-X.
12. Kargin A. A. Introduction to intelligent machines. Book 1. Intelligent controllers [text] / A. A. Kargin. – Donetsk: Nord-Press, DonNU, 2010. – 526 p.
13. Kargin A. A. Learning models of motivated context situational control system [text] / A. A. Kargin, N. V. Krachkovsky – Kherson: KhNTU Bulletin – 2012.-№1(44).-P.257–260

RESUME

M.V. Krachkovsky

About Reinforcement Learning Method Modification Based on Cognitive Psychology Models

In this article we consider the problem of learning of motivated context situational control system of the complex system behavior. The structural changes occurring in the control system, represented as a set of situational agents at training, are considered. Developed conceptual model is based on the researches of the physiologists and cognitive psychologists.

The formal learning model is developed, which describes the changes in the situation and reaction prototypes and the various parameters influence on learning, such as the amount of reinforcement, the time between the stimulus, action and the reinforcement. There is also described the change of the contextual link between situational elements when it's used.

The training consisting in formation of the new situational agent which includes formation of two situational elements from the empty unallocated element is described. Results of experiments of computer modeling of training are shown.

Статья поступила в редакцию 10.06.2013.